Washington University School of Medicine
Digital Commons@Becker

Independent Studies and Capstones

Program in Audiology and Communication Sciences

2012

Gaze stabilization testing: The efficiency and test-retest reliability of user operated control versus the parameter estimation (PEST) algorithm

Michelle E. Hungerford
Washington University School of Medicine in St. Louis

GAZE STABILIZATION TESTING: THE EFFICIENCY AND TEST-RETEST RELIABILITY OF USER OPERATED CONTROL VERSUS THE PARAMETER ESTIMATION (PEST) ALGORITHM

by

Michelle E. Hungerford

A Capstone Project
submitted in partial fulfillment of the
requirements for the degree of:

Doctor of Audiology

Washington University School of Medicine
Program in Audiology and Communication Sciences

May 17, 2013

Approved by:
Belinda Sinks, Au.D., Joel Goebel, M.D., Capstone Project Advisors
Maureen Valente, Ph.D., Second Reader

Abstract: The primary objective of this research study is to determine which form of testing, the PEST algorithm or an operator-controlled condition is most accurate and time efficient for administration of the gaze stabilization test

ACKNOWLEDGMENTS

I would like to extend thanks to Dr. Belinda Sinks for her tireless support, insight and guidance throughout the completion of the capstone project. I would also like to thank Dr. Joel Goebel for his support, insight and guidance as a second mentor. I would also like to thank Dr. L Maureen Valente for her valuable contributions as a collaborator and second reader. Melissa Krauss, MPH of the Institute of Clinical and Translational Sciences Research Design and Biostatistics Group in the Division of Biostatistics at Washington University School of Medicine provided invaluable assistance with statistical analysis and data interpretation. I also thank Dr. Jon Peters for his thoughtful suggestions. I would like to express sincere appreciation for the participants who dedicated their time to this study. Finally, I would like to thank my family for their unwavering support and instilling in me a passion for learning.

This publication was made possible by Grant Number UL1 RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

DIE	Dynamic illegible "E"
DVAT	Dynamic visual acuity test
GST	Gaze stabilization test
ICC	Intra-class correlation coefficients (ICC)
logMAR	Logarithm of the Minimum Angle of Resolution
PEST	Parameter estimation by sequential testing
SVA	Static visual acuity
VOR	Vestibulo-ocular reflex
VPT	Vestibular physical therapy
WVC	Within-subject coefficient of variation

INTRODUCTION

A person's balance is influenced by the interaction of the ocular-motor, proprioceptive, and vestibular systems. The vestibulo-ocular reflex (VOR) acts as a mode of compensation during head movement that aids in the stabilization of vision. As the head rotates, the VOR generates compensatory eye movements that allow for clear vision during movement. This system is especially important at high velocities and frequencies of head motion, such as during walking or running.

When the vestibular system is compromised and is not functioning properly, some patients experience oscillopsia, which is a blurring of vision, while the head is in motion (Gresty, M. A., Hess, K., & Leech, J., 1977; Chambers, B. R., Mai, M., & Barber, H. O., 1985; Bhansali, S. A., Stockwell, C. W., & Bojrab, D. I., 1993). This can lead to decreased quality of life for those affected. Assessing the VOR allows clinicians to look at the impact of the vestibular loss on the patient's function, as well as assess the effectiveness of interventions for these patients. Two of the clinical tests that have come to fruition in order to assess VOR function during high frequency head movements are the dynamic visual acuity test (DVAT) and the gaze stabilization test (GST). The DVAT and the GST both allow a clinician to see functional impairment in the VOR. The GST has been studied in the past few years and has proven to be a useful upcoming clinical device for testing the VOR.

The first clinical forms of VOR testing were introduced in the 1980's. One of the first forms of VOR testing was the head-thrust test, which was introduced in 1988 by Halmagyi and Curthoys. For this test, the patient fixates on an object while the physician moves the patient's head quickly first in one direction and then the other. If the VOR is functioning normally, the patient should be able to keep his/her focus on the object. Another form of this test involves a patient reading a standard Snellen eye chart with his/her head static, and then again in motion (dynamic condition). Results are obtained by looking at the difference between the static and dynamic measures obtained. If significant differences in the patient's ability to read the eye chart are noted, it is considered an indication of VOR impairment.

From these preliminary tests rose the need for a test with better accuracy, which then led to the development of the clinical DVAT. This test began as the dynamic illegible "E" (DIE) test, which used the letter "E" in place of the Snellen eye chart. This was implemented to reduce correct answers due to memory of the chart or incorrect answers due to the unequal legibility of the letters (Longridge, N. S., & Mallinson, A. I., 1987). Although this test correlated with another test within the battery (bithermal caloric irrigations) regarding functioning of the patient, the DIE could not monitor the patient's head movement to determine whether testing was accurate.

In an effort to deviate from the uncontrolled test conditions of the bedside exams mentioned above, the computerized DVAT was developed. In the DVAT, patients were asked to shake their heads from side to side as if saying "no" at a constant speed while looking at a computer screen. The ototype "E" was then presented to the patient in either larger or smaller fashion depending on whether the patient had identified the previous ototype correctly. For example, if the previous ototype was not identified correctly, the ototype became larger during the next presentation. In similar fashion, a smaller ototype appeared if a correct response was obtained on the previous presentation. The test was problematic in that the differing sizes of the ototype resulted in false negative responses that were due to visual issues, as opposed to true vestibular problems (Goebel, J. A., Tungsiripat, N., Sinks, B., & Carmody, J., 2006).

The GST, which was the next test in the evolution of the high frequency VOR assessments, was developed to remedy the crossover from the visual system that can be seen in the DVAT. Instead of seeing a change in size, the patients must shake their heads at different speeds while the ototype size remains the same. By keeping the ototype size constant, the GST test is less susceptible to false positive responses due to visual impairments. The GST has been highly researched in current literature (Pritcher, M. R., Whitney, S. L., Marchetti, G. F., & Furman, J. M., 2008; Goebel et al., 2006; Ward, B. K., Mohammad, M. T., Whitney, S. L., Marchetti, G. F., & Furman, J. M., 2010; Whitney, S. L., Marchetti, G. F., Pritcher, M., & Furman, J. M., 2009; Honaker & Shepard, 2010; Gottshall & Hoffer, 2010; Gottshall, 2011) in order to assess its viability as a clinically applicable instrument. A compact version of the GST has been introduced to help to further standardize testing. The tunnel system of the GST utilizes mirrors that reflect the image of a computer monitor in order to achieve a constant distance between the patient and the target. Figure 1 shows a patient sitting at the GST tunnel system facing the mirror, which reflects the computer screen. A tracker is placed on the patient's head in order to record head velocity. The clinician sits at the computer located behind the patient to administer the test.

Many different aspects of GST have been studied thoroughly in recent literature in order to justify its future use as a widely utilized clinical test. One of the main research interests has been to identify the ability of the GST to be sensitive to vestibular dysfunction. For patients with vestibular dysfunction, the GST results indicate that these patients have a slower head velocity, or that they are not able to move their heads from side to side as quickly as those without vestibular dysfunction (Pritcher, 2008; Goebel, 2006). In a study by Goebel et al. (2006), results showed that patients with unilateral vestibular dysfunction could be distinguished from the normal control group with fair sensitivity and high specificity. This study also demonstrated that VOR function was reduced on the affected as well as the unaffected side. In a study by Pritcher et al. (2008), patients with many different types of vestibular dysfunction were tested and results concluded that the vestibular impaired group as a whole showed slower velocities than the control groups.

Another area of study regarding the GST has been the effect of age on head velocity. Due to the demands of the testing setup that primarily relate to head and neck mobility, it is possible that older patients may be more susceptible to false negative responses. Results reported in current literature, however, have been variable. Pritcher et al. (2008) reported finding no significant differences in head velocity between older and younger control subject groups. In a study by Honaker et al. (2010), however, age was reported as a significant factor in GST velocity. This study reported that maximum head velocity decreased as age increased. Studies by Whitney et al. (2009) and Ward, B. K., Mohammed, M. T., Brach, J. S., Studenski, S. A., Whitney, S. L., & Furman, J. M. (2010) also showed that GST velocity decreased with increasing age. Differing subject inclusion criteria and methodology, such as longer ototype duration, differing health status, and more advanced age may have contributed to variability on study outcomes. As standardization becomes more readily available for this test, the effect of age may become clearer and more standardized as well.

Other researchers are utilizing the GST as a tool to gauge progress in vestibular physical therapy (VPT) (Gottshall & Hoffer, 2010; Gottshall, 2011). This line of investigation is especially relevant to vestibular deficits secondary to traumatic brain injury. As this is the most common wound found in combat, researchers are looking for a way to assess vestibular function in order to return soldiers to the field (Gottshall & Hoffer, 2010). When presented in a test battery approach to track VPT progress, Gottshall & Hoffer (2010) found that the GST results did not return to normal levels until 12 weeks of VPT were completed. The authors also stated that soldiers able to run three miles without symptoms at this 12-week evaluation were those who also had normal GST values at that time. In a study by Gottshall (2011), the results suggest that the GST in the pitch plane was the most sensitive indicator of patient outcomes in a population of young soldiers with mild traumatic brain injury. In this kind of approach, the GST may be useful in determining the functional gains from VPT.

In a study by Ward et al. (2010) the GST's reliability, validity and stability were tested for the tunnel system. This study demonstrated that in a control group, the GST is reliable and measures the VOR function. When looking at test-retest reliability, the GST had excellent reliability within the same test session and also between test sessions. The GST was also found to be more reliable in the pitch (vertical) plane for the younger adults than for older adults. This may be due to the constant quick head movements required to accurately complete the GST, thus causing a greater sense of fatigue in older subjects.

Ongoing research in the area of clinical utilization of the GST relates to how the test itself is scored. The GST utilizes a modified parameter estimation by sequential testing algorithm (PEST), which is an algorithm that measures psychophysical thresholds. These thresholds involve the relationships between physical and sensory inputs. PEST was developed by Taylor and Creelman with the intent to change the level of the stimulus, or entity that is used to elicit a response (Taylor, M. M., Forbes, S. M., & Creelman, C. D., 1983). The PEST was also intended to change step size, or the increment of change, to quickly and efficiently determine the targeted level of performance. The original PEST algorithm determines threshold, or the psychophysical parameter that is desired, by utilizing a modified staircase procedure. A staircase procedure is one in which the stimuli are separated by an interval, which is referred to as the step-size, and are presented in either ascending or descending order until the participant's threshold is found. Staircase-like measures are often used to determine different physical measures to determine a threshold in the quickest and most accurate way while gaining the maximum amount of information from each trial (Liberman & Pentland, 1982). During the creation of the procedure, the staircase method did not involve utilizing the information obtained on the previous trials. The PEST algorithm evolved in the 1960's and the 1970's to include the use of information from previous trials, thus reducing the number of trials that were necessary to reach threshold using this algorithm. A PEST algorithm will either double or halve the stimulus level based upon the previous responses collected. This way, the PEST can assess if threshold has been exceeded during each trial performed (Gelfand, 2009).

A modified PEST algorithm was derived from the original PEST by Pentland in 1980 and was named the Best PEST. The Best PEST utilizes a maximum likelihood estimation, which estimates the parameters of a statistical model, in order to reduce the number of trials needed to establish threshold and provide the maximum information possible regarding the threshold level. The Best PEST has also been modified for the purposes of the GST. This modified PEST algorithm is incorporated into the Dynamic Vision Software that is utilized during the GST. One challenge of the Best PEST testing was that it required too many trials for the GST in order to achieve a relevant threshold. The termination criterion of the test was modified to allow for fewer trials while maintaining accuracy, thus giving way to the modified PEST protocol. After each level is tested with the modified PEST, the computer assesses to determine if the end criteria are met for that level and for a level to be accepted as threshold. There must be at least three trials performed and an overall positive response of ≥ 60%. The level that is one step more difficult than the level accepted as threshold must also have at least three trials performed and an overall negative response ≥ 60%. If all of the criteria for both levels are met, then the level with the overall response ≥ 60% can be named as threshold once the results can be replicated twice more at the same level. The other major change from the Best PEST to the modified PEST is that the programming has been changed to reflect that the GST uses a four choice closed set task. For the GST, the participant is asked to tell whether the "E" is facing up, down, right, or left. This modified PEST algorithm is the current standard testing protocol for the GST.

In the case of the GST, VOR threshold is the psychophysical parameter that is desired, which is the maximum head velocity in which the participants can reliably detect the stimulus. During testing under the PEST criterion, positive responses are followed by an increase in the stimulus level in order to become closer to threshold, and negative responses are followed by a decrease in the stimulus level. Threshold is characterized by a change in the participant's answers because it is the level where the participant is no longer consistently responding correctly. Threshold for the GST is the level in which the participants can distinguish three out of five stimuli correctly.

Another clinical measure that is available through use of the GST In-Vision software is to test utilizing an operator-controlled condition. In this testing paradigm, the clinician is able to choose a starting velocity and move upward or downward in velocity as he/she sees fit in order to achieve threshold. This procedure is very much like the PEST procedure in that it entails ascending or descending in set increments that are based upon the participant's response. The major difference between the PEST condition and the operator-controlled condition is that the operator-controlled condition has the ability to take outside factors into account. For example, if a participant happens to blink as the ototype is presented and is unable to identify its direction, or loses attention, it is possible for the clinician to count the trial as invalid and re-test that trial. In this situation, the modified PEST algorithm would assume that the participant was unable to complete the trial because the velocity had exceeded his/her threshold, which is not the case. The modified PEST is not able to take into account participant error, whereas this can be accounted for in the operator controlled setting.

One of the main challenges that still remains in using the GST clinically is the way the test is scored by the modified PEST algorithm. Although the modified PEST is a valid psychophysical tool, it may not be the most accurate for the current GST testing protocol. During the testing session, the computer is unable to determine if the patient is experiencing a true dysfunction of the VOR when trials are failed, or if these trial failures are due to the inability of the patient to move his/her head quickly enough (Ward, 2010). One way to ensure that the test session has the least amount of false negative responses possible is to allow the testing conditions to be conducted by a clinician.

The aim of this study was to compare test results of clinician versus computer driven procedures. It is hypothesized that an operator-controlled testing paradigm may yield more accurate results and be more time efficient for administration of the gaze stabilization test than the PEST algorithm.

METHODS

Participants

The tunnel system gaze stabilization testing was performed on healthy adult volunteers from Washington University in St. Louis and the surrounding St. Louis community. The participants taking part in this study gave written informed consent prior to participation. There were ten females and ten males in the study, ranging from 19 to 36 years of age. The mean age of participants is 25 years with a standard deviation of 3.13 years. All of the participants reported normal vestibular function, normal or corrected eyesight, and absence of any head or neck conditions that would be contraindicative to fast head movement. Participants were not paid for their participation. The Human Research Protection Office at Washington University in St. Louis School of Medicine (WUSM) approved this study.

Testing Procedure

The participants were evaluated with the tunnel gaze stabilization test with all testing performed within the WUSM Dizziness and Balance Center. They were situated in a chair directly in front of the system facing the computer screen in a darkened room. The Neurocom tunnel system utilizes mirrors to ensure that the participant is always a consistent four meters from the computer screen when seated in front of the device. Testing occurred in two separate sessions lasting approximately sixty minutes each. These test sessions were at least one day apart, depending on participant availability. Each test session included: static visual acuity, perception time, and four GST trials, two with the modified PEST algorithm and two that were operator controlled.

Participants were first tested for static visual acuity. For static visual acuity, the participants sat in front of the computer screen with their heads stationary. They were instructed to report the orientation of the optotype "E" (up, down, right, left). The participant was instructed to respond with "I don't know" if unable to choose. The ototype became smaller with every trial until the computer determined the smallest "E" which the participant could correctly identify in 3 out of 5 presentations. The ototypes were presented .25 logMAR, or 2.5 lines on a Snellen eye chart, above the static visual acuity score obtained from each participant. Both the computer PEST algorithm as well as the user operated testing conditions used this set interval.

Participants were then tested for their perception time. In this test, the participants were presented the ototype "E" at .25 LogMAR above their static acuity score. The computer software assessed the shortest time (ms) the participant needed to correctly identify the orientation of the ototype. The value obtained from this test was used in the GST to set the duration of the presentation of the ototypes. If the participant's minimum perception time was less than 40 milliseconds, then the minimum presentation time was set at 40 milliseconds and the maximum at 75 milliseconds. Due to the control population used for this study, the minimum perception time was always under 40 milliseconds and thus testing parameters were set at the same value for each participant.

Once the parameters of the test were set, a tracker was placed on top of the participant's head to measure head velocity, and then he/she was provided with instructions for the GST test. Participants were instructed to shake their heads from side to side in the yaw plane while looking at a target in the middle of the computer screen. They were asked to keep their head movements such that the indicator on the screen remained green. This indicated that the head was being kept at a constant speed within the target range. Participants were allowed to practice both head movements and the GST test itself until they reported comfortable with the task. Once the participant was competent with the task, the GST was performed. The GST tests different head velocities in order to obtain an average of the three fastest head speeds of a participant while he/she was able to correctly identify the direction of the ototype. If the correct head velocity was achieved for a trial, the "E" ototype would be shown on the screen. If the head velocity was not correct for a trial, it was marked as a failed attempt and the ototype was not displayed on the screen. Throughout testing, participants were allowed rest periods to reduce fatigue. The rest periods were given as needed, as well as between the two testing algorithms and between each trial.

For the purposes of this study, a Hughson-Westlake procedure, which is widely used in the audiology profession to collect audiometric data, was utilized to determine threshold during the operator-controlled condition. In the operator-controlled condition, head velocity was increased in increments of twenty degrees per second until the participant could no longer correctly identify the direction of the ototype. Once this level was achieved, the level was dropped by 10 degrees per second until a correct response was elicited. Once a correct response was elicited, the level was increased by 5 degrees per second until the highest level with two correct responses was found. This level was recorded as threshold. These trials were conducted in randomized order over both test sessions.

RESULTS

The purpose of the present study was to examine the effects of testing programs on the GST and determine which is most accurate and time-efficient. Results were obtained utilizing two different testing methods (PEST and operator-controlled.) Scores from each testing method and trial were recorded as well as the time it took to administer the test. A repeated measures analysis of variance (ANOVA) was completed to examine whether the mean time to complete the test and the test scores differed under PEST and operator-controlled algorithms and between the first and second trials within a session. Test-retest reliability for scores from the gaze stabilization test were determined for each algorithm type separately (operator-controlled and PEST) using intra-class correlation coefficients (ICCs) with 95% confidence intervals. Within-subject coefficient of variation (WCV) was also calculated with 95% confidence intervals. ICCs and WCVs were determined for both sessions combined, for repeat trials within each session, and for the corresponding trial across sessions.

Time

A repeated measures analysis revealed a significant difference in the administration time of the GST between testing methods. Time to administer the test was greater for the operator-controlled paradigm than for the PEST. The mean difference between the testing conditions for the first session was 4.4 minutes ($F(1,57)=83.6$, $p<.001$) and 3.1 minutes ($F(1,56)=89.0$, $p<.001$) for the second session. There was a significant main effect for session ($F(1,132)=4.1$, $p=.049$) and a significant interaction between session and type ($F(1,132)=7.7$, $p=.006$). Upon examination with post-hoc tests (and applying Tukey HSD for multiple comparisons), there was no difference in time between Session 1 and Session 2 for PEST (mean difference -0.2 minutes, t(132)=-0.5, p=.951), but time was greater in Session 1 versus Session 2 for the USER algorithm (mean difference 1.1 minutes, t(132)=3.4, p=.005). Figure 2 demonstrates this difference in time between both testing conditions as well as between sessions.

Test-Retest Reliability

Test-retest reliability was determined for each testing condition separately using intra-class correlation coefficients (ICC) and within-subject coefficient of variation (WVC) with 95% confidence intervals. Test-retest reliability did not reach statistical significance due to wide and overlapping confidence intervals between the testing conditions. Table 1 contains the ICC and WVC for overall scores, within session scores, and also scores across sessions for each of the trials. Another way to examine the correlations within the testing conditions is to view the scatter of scores. Figure 3 demonstrates the scatter seen within each condition across sessions and trials.

Head Velocity

Scores on both the right and left sides were significantly greater for the operator-controlled paradigm compared to the PEST algorithm when compared with a two-way ANOVA (mean difference on right side 70.4 (F(1,57)= 33.4, p<.001) and mean difference on left side 79.3 (F(1,57)= 37.52, p<.001)). Statistically significant differences were also seen for the second trial than the first trial (mean difference on right side 20.2 (F(1,57)= 49.88, p=.024) and mean difference on left side 39.7 (F(1,57)= 26.51 p<.001)). Figure 4 shows the average maximum head velocities achieved across patients for all testing conditions, trials, and sessions. As Figure 4 demonstrates, the operator-controlled paradigm was able to elicit larger maximum head velocities than the PEST algorithm across both trials and sessions. The operator-controlled condition elicited the largest maximum head velocities during the second trial of the second session.

Although the operator-controlled condition elicited larger maximum head velocities, maximum head velocities tended to become larger in the later trials of both testing conditions. As figure 4 demonstrates, the maximum head velocity scores, in general, were higher than those of the previous trial with a notable difference between the first and second sessions. Overall, scores on the right were significantly higher during Session 2 than Session 1 ($F(1,133)=8.7$, mean difference 21.0, $p=.004$). This was also the case for scores on the left ($F(1,133)=15.2$, mean difference 29.6, $p<.001$).

Differences Between Left and Right Sides

Although this study utilized healthy adults as the test population, there were some differences in maximum head velocities between the left and right sides. The velocities obtained to the left side were significantly different between the trials of both testing conditions of the first testing session ($F(1,57)= 37.52$, $p<.001$). This effect did not exist during the second session or within velocities obtained from the right side.

DISCUSSION

The current study utilized multiple testing paradigms to elicit thresholds of the GST. Participants were all healthy adults who did not report the presence of any vestibular dysfunction and thus were expected to have normal maximum head velocity thresholds. Results obtained suggest that there are differences between the two testing conditions and that the operator-controlled condition may be more reliable and reproducible.

Effects of Testing Time

Although the finding of prolonged testing time is contrary to our hypothesis, it can be explained through the study design. One of the possible reasons for this discrepancy is that the PEST was utilized at its default parameters. These parameters allow the algorithm to test between 20 and 300 degrees per second. The operator-controlled paradigm did not work within the limits of an algorithm and the operator was able to test above 300 degrees per second. The goal of this study was to find threshold in normal subjects instead of screening for intact VOR function for patients with suspected vestibular dysfunction. As a result of the search for the best performance velocity, the operator was able to administer the test at higher levels than the computer PEST algorithm was programmed to test. Part of this time discrepancy is possibly due to higher testing velocities and more difficult test conditions.

This study examined the highest maximum head velocities that could be elicited from the participants. When the GST is used in a clinical setting, it is often used via a screening mode and not its default parameters. The screening mode available utilizes the PEST algorithm with a maximum of 150 degrees per second, and velocity obtained above this level would indicate normal vestibular function. If there was a cut off velocity, such as 150 degrees per second, it is possible that operator-controlled paradigm may be administered more quickly than the PEST. Further study is needed to examine the relationship between the testing conditions in a screening type of test application.

This study demonstrated that the time taken to administer the operator-controlled paradigm was statistically significant between sessions, whereas the PEST was not. One explanation for the difference is a learning effect on the part of the clinician. As the clinician became more proficient with the testing conditions, it is possible that she was able to perform the test faster during the second session.

Test-Retest Reliability

The current study found a lower correlation of test-retest reliability than reported in previous studies (Goebel et al., 2006; Ward et al., 2010). Although the ICC's are low, they are better for the operator-controlled paradigm than the PEST algorithm. ICC's were also lower between sessions than within sessions, suggesting a weaker correlation between testing sessions. WCV's, like the ICC's, are better for the operator-controlled condition as well. Also, figure 3 demonstrates that the PEST algorithm has more scatter of maximum head velocity scores than the operator-controlled paradigm. The body of data suggests that there is a trend toward the operator-controlled condition being more reliable and having better reproducibility than the PEST algorithm both within and between sessions in this study. However, this study did not have a large enough sample size to reach statistical significance for this trend.

One possible explanation for the variability in the test-retest reliability seen in this study is the variability in static visual acuity (SVA) scores within participants. Any change in a participant's SVA will affect the difficulty of the test because it will change the size of the ototype seen during testing. Small SVA scores result in small ototypes during the GST, thus making the test more difficult. Fourteen of the twenty participants in the study had a different SVA score for the second session than they did in the first session. It is possible that this level of variation in the static visual acuity scores may be larger than in other studies and may contribute to some of the variability seen in the data.

Another possible area that may contribute to the lower test-retest reliability seen in this study is the amount of practice time given to each participant. It is possible that participants received less practice before the testing procedure began than in previous studies. The clinician began practice headshakes at 100 degrees/second and ended with 140 degrees per second using a step increase of 20 degrees per second. The participant was asked to perform practice headshakes until he/she felt comfortable with the task and was able to perform all practice velocities correctly. It is possible that previous testing included a more diverse or rigorous practice session, thus resulting in more stable scores.

Within this study, one issue that was seen with the PEST algorithm that was not present in the operator-controlled paradigm was that the PEST condition had to be re-tested on six separate trials due to maximum head velocities that indicated vestibular dysfunction. The ability of healthy participants to fail trials is a challenge that has also been documented in previous research (Ward et al., 2010). There were no trials that needed to be re-tested for the operator-controlled condition, as the operator was able to control for trial failure due to outside variables by repeating the defective head velocities and disregarding the corrupted trial.

Effects of Maximum Head Velocity

This study found larger maximum head velocities than previous studies (Goebel et al, 2006; Ward et al., 2010). The presence of these large velocities may be due to testing for the absolute highest threshold for each participant. The default GST software allowed the PEST to test up to 300 degrees per second and the clinician was able to test even higher head velocities. Participants were driven to their highest possible head velocities utilizing the large range of velocities available for testing.

The maximum head velocities of the operator-controlled paradigm were consistently larger than those of the PEST algorithm. This finding may be due to limitations of the PEST algorithm to control for failed testing velocities resulting from factors other than vestibular dysfunction. The PEST is unable to distinguish whether a participant failed a head velocity due to vestibular dysfunction or other factors, such as lack of head motion or an interruption, such as a sneeze. The PEST may have been rejecting velocities that could have been valid if additional testing was completed. Another factor in this score discrepancy is that the clinician was able to test at velocities above the maximum cutoff for the PEST algorithm to reach threshold. Although the PEST algorithm includes ranges that test for vestibular dysfunction, it may be underestimating a participant's absolute threshold.

Another finding in this study suggests that velocities tended to become larger in the later trials of both testing conditions. Figure 4 demonstrates that there seems to be a learning effect that occurs during the GST. Scores were statistically significantly higher during the second session than the first session for both right and left sides across both testing conditions. This finding supports a learning effect that takes place with this test regardless of which testing paradigm is utilized.

Effects of Directions Tested

The current study found that there was a significant effect of direction tested during the first session. During this first session, the data obtained from the left side may be worse than data from the right side because the PEST algorithm defaults to begin testing with the left side. Because of this testing parameter, the operator-controlled condition also began with the left side as a default. Due to the fact that the left side was always tested first, it is possible that this effect is a result of the participants' inexperience with the test. As the participants became more familiar with the task, no statistically significant differences between the left and right data were found for the second session.

Future Directions

The current study generated intriguing findings as well as questions for further study. The study could be repeated with a larger sample size in order to achieve statistical significance and clarify the differences between groups. Another direction of study is to document the relationship between the PEST and operator-controlled paradigms utilizing a screening mode. The screening mode offers a quick and efficient assessment to rule out vestibular dysfunction. This mode is currently available on the GST software and utilizes a cutoff of 150 degrees per second for the PEST algorithm. Comparing the PEST and operator-controlled conditions within the screening mode may offer insight into a faster and more reliable screening test.

One more direction of study is to test the reliability of the operator-controlled paradigm in a patient population. Although this study discovered a trend that suggests the operator-controlled paradigm might be more reliable and reproducible, it is essential that its use be studied in a patient population to ensure that it is a reliable measure to clarify the presence of vestibular dysfunction. The operator-controlled paradigm has proven to be a promising testing technique, but further research is needed to better understand its function.

Lastly, another direction of study is the effects of the operator-controlled condition when the clinician is not an audiologist. Audiologists are familiar with a threshold search procedure since many of the clinical audiological tests rely on this form data analysis. The success of the operator-controlled condition may or may not be contingent upon the operator's professional skills. This line of research may include the validity and reliability of the paradigm across different professions.

CONCLUSIONS

This study revealed significant differences between the testing conditions (PEST and operator-controlled) of the GST. Results demonstrate that the operator-controlled condition was able to control for failed testing velocities due to factors other than vestibular dysfunction, whereas the PEST was not able to distinguish the difference. This study demonstrated that the operator-controlled condition did take longer to administer than the PEST, but there is a trend that suggests it may be more reliable and reproducible. Further testing is needed to verify the use of the operator-controlled paradigm in clinical gaze stabilization testing.

REFERENCES

Bhansali, S. A., Stockwell, C. W., & Bojrab, D. I. (1993). Oscillopsia in patients with loss of vestibular function. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *109*(1), 120–125.

Chambers, B. R., Mai, M., & Barber, H. O. (1985). Bilateral vestibular loss, oscillopsia, and the cervico-ocular reflex. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *93*(3), 403–407.

Gelfand, S. (2009). Psychoacoustic Methods. *Hearing : An Introduction to Psychological and Physiological Acoustics (5th Edition, Revised and Expanded)* (pp. 146–159). New York, NY, USA: Informa Healthcare. Retrieved from http://site.ebrary.com.beckerproxy.wustl.edu/lib/beckermed/docDetail.action?docID=10369928

Goebel, J. A., Tungsiripat, N., Sinks, B., & Carmody, J. (2007). Gaze stabilization test: a new clinical test of unilateral vestibular dysfunction. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *28*(1), 68–73.

Gottshall, K. (2011). Vestibular rehabilitation after mild traumatic brain injury with vestibular pathology. *NeuroRehabilitation*, *29*(2), 167–171. doi:10.3233/NRE-2011-0691

Gottshall, K. R., & Hoffer, M. E. (2010). Tracking recovery of vestibular function in individuals with blast-induced head trauma using vestibular-visual-cognitive interaction tests. *Journal of Neurologic Physical Therapy: JNPT*, *34*(2), 94–97.

Gresty, M. A., Hess, K., & Leech, J. (1977). Disorders of the vestibulo-ocular reflex producing oscillopsia and mechanisms compensating for loss of labyrinthine function. *Brain: A Journal of Neurology*, *100*(4), 693–716.

Halmagyi, G. M., & Curthoys, I. S. (1988). A clinical sign of canal paresis. *Archives of Neurology*, *45*(7), 737–739.

Honaker, J. A., & Shepard, N. T. (2010). Age effect on the Gaze Stabilization test. *Journal of Vestibular Research: Equilibrium & Orientation*, *20*(5), 357–362.

Lieberman, H., & Pentland, A. (1982). Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behavior Research Methods*, *14*(1), 21–25.

Longridge, N. S., & Mallinson, A. I. (1987). The Dynamic Illegible E-test: a technique for assessing the vestibulo-ocular reflex. *Acta Oto-Laryngologica*, *103*(5-6), 273–279.

Pritcher, M. R., Whitney, S. L., Marchetti, G. F., & Furman, J. M. (2008). The influence of age and vestibular disorders on gaze stabilization: a pilot study. *Otology & Neurotology:*

Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, *29*(7), 982–988.

Taylor, M. M., Forbes, S. M., & Creelman, C. D. (1983). PEST reduces bias in forced choice psychophysics. *The Journal of the Acoustical Society of America*, *74*(5), 1367–1374.

Ward, B. K., Mohammad, M. T., Whitney, S. L., Marchetti, G. F., & Furman, J. M. (2010). The reliability, stability, and concurrent validity of a test of gaze stabilization. *Journal of Vestibular Research: Equilibrium & Orientation*, *20*(5), 363–372.

Ward, B. K., Mohammed, M. T., Brach, J. S., Studenski, S. A., Whitney, S. L., & Furman, J. M. (2010). Physical performance and a test of gaze stabilization in older adults. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *31*(1), 168–172.

Whitney, S. L., Marchetti, G. F., Pritcher, M., & Furman, J. M. (2009). Gaze stabilization and gait performance in vestibular dysfunction. *Gait & Posture*, *29*(2), 194–198.

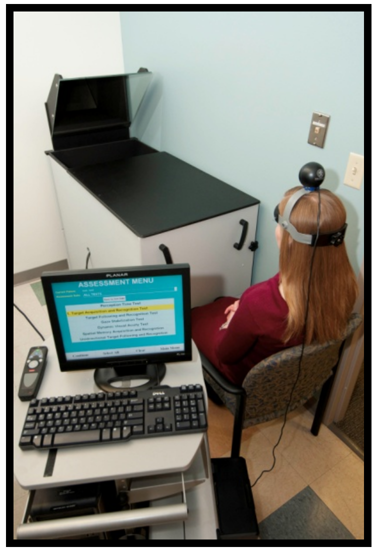


Figure 1. *The tunnel system GST testing equipment (Neurocom).*

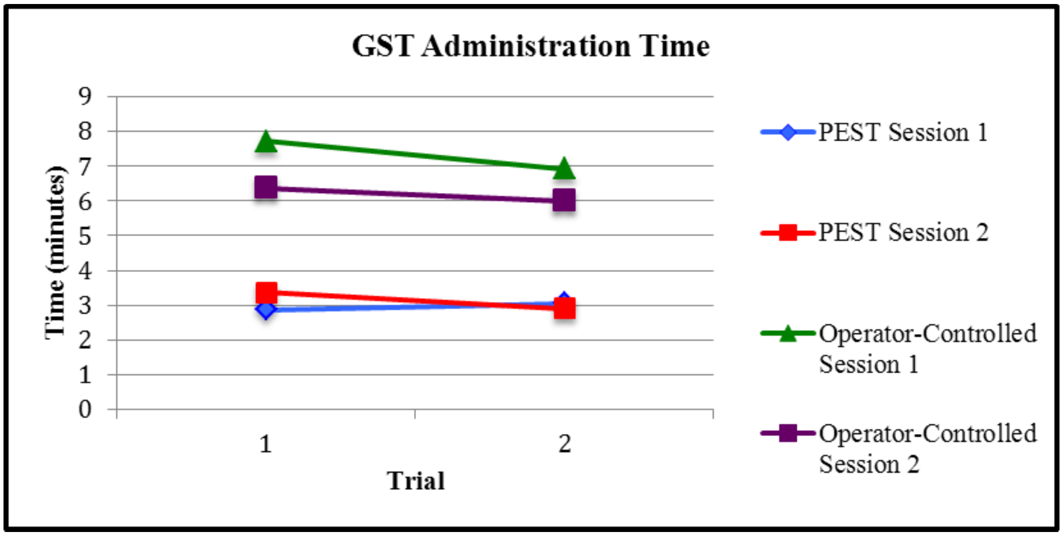
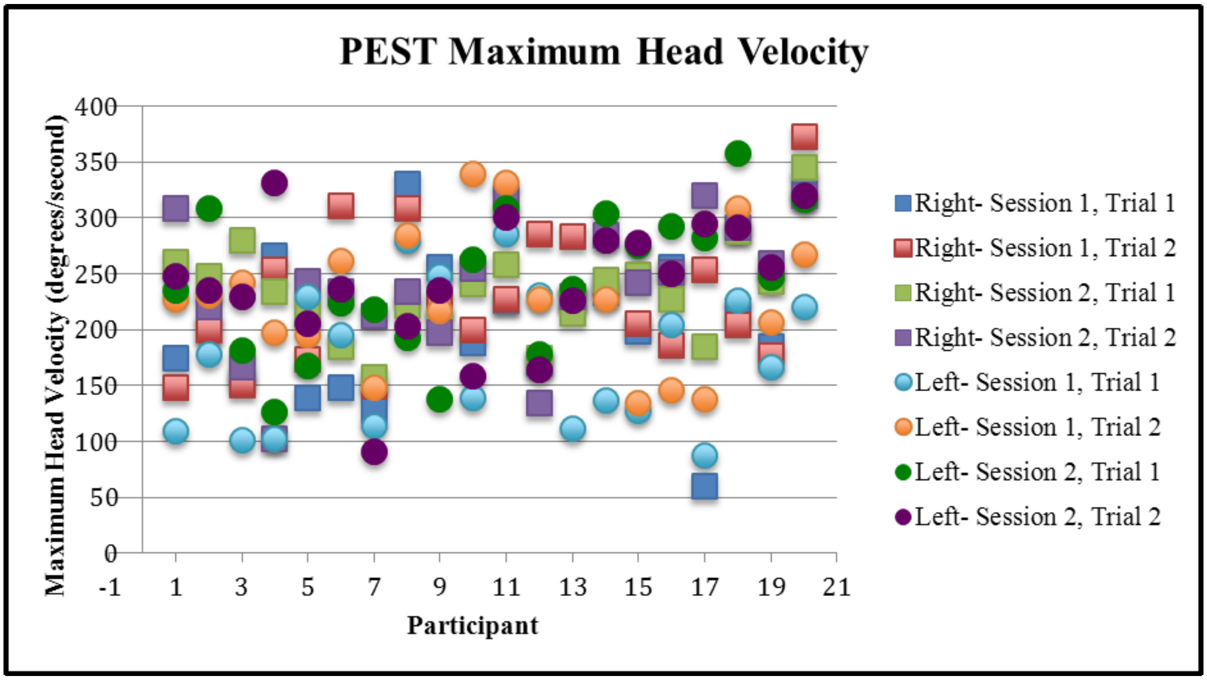


Fig. 2. *Average mean times to administer the test by algorithm (Operator and PEST) and session*

Table 1. Intra-class correlation coefficients and within-subject coefficient of variation by algorithm and session.

	PEST		USER	
	ICC (95% CI)	WCV (95% CI)	ICC (95% CI)	WCV (95% CI)
Overall				
Score on right	0.19 (0.05 – 0.51)	0.23 (0.19 – 0.28)	0.44 (0.24 – 0.67)	0.11 (0.10 – 0.14)
Score on left	0.10 (0.01 – 0.54)	0.28 (0.23 – 0.34)	0.38 (0.19 – 0.63)	0.13 (0.11 – 0.16)
Within session 1				
Score on right	0.42 (0.14 – 0.76)	0.22 (0.16 – 0.31)	0.49 (0.20 – 0.79)	0.11 (0.08 – 0.15)
Score on left	0.22 (0.02 – 0.76)	0.29 (0.21 – 0.40)	0.43 (0.15 – 0.77)	0.12 (0.09 – 0.16)
Within session 2				
Score on right	0.35 (0.09 – 0.75)	0.17 (0.13 – 0.24)	0.75 (0.52 – 0.89)	0.07 (0.08 – 0.15)
Score on left	0.34 (0.08 – 0.74)	0.20 (0.15 – 0.28)	0.69 (0.43 – 0.87)	0.09 (0.07 – 0.13)
Across sessions				
1st trial				
Score on right	0.37 (0.10 – 0.75)	0.20 (0.15 – 0.28)	0.32 (0.07 – 0.74)	0.13 (0.10 – 0.18)
Score on left	0	0.34 (0.27 – 0.43)	0.13 (0.00 – 0.87)	0.14 (0.10 – 0.19)
2st trial				
Score on right	0.03 (0.00 – 1.00)	0.25 (0.18 – 0.34)	0.34 (0.09 – 0.74)	0.12 (0.09 – 0.16)
Score on left	0.02 (0.00 – 1.00)	0.24 (0.18 – 0.34)	0.36 (0.09 – 0.75)	0.14 (0.10 – 0.19)



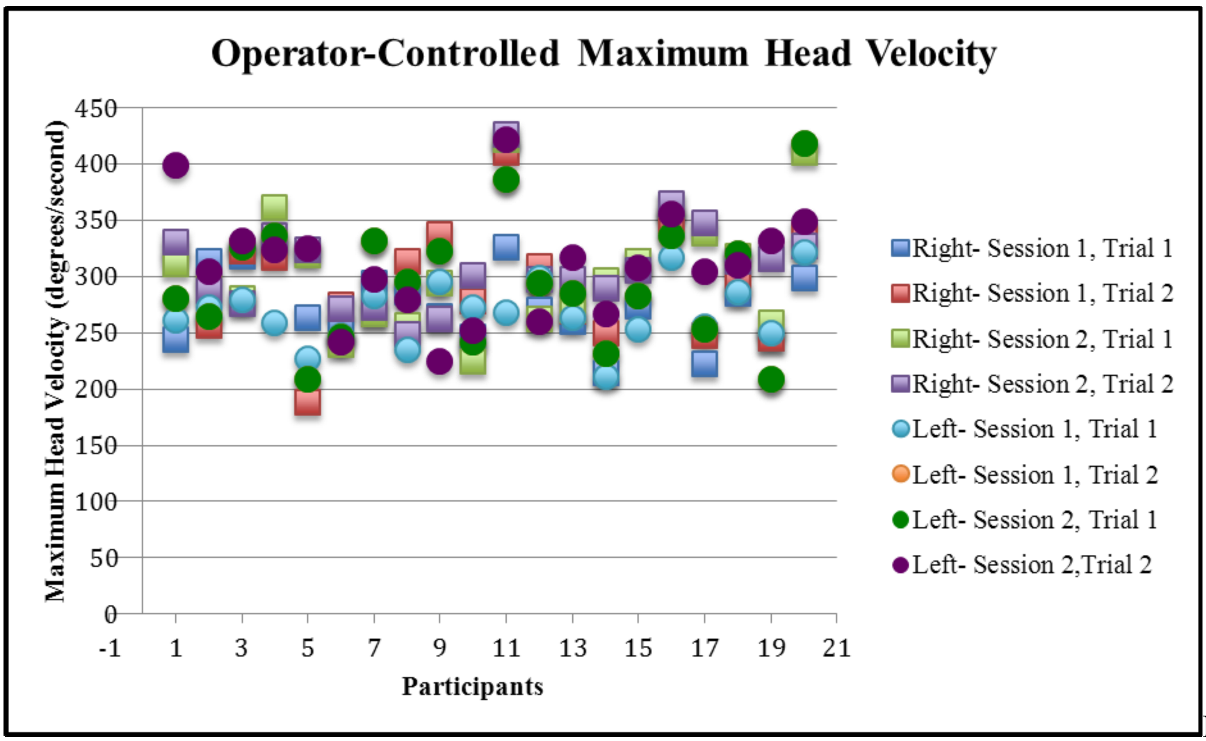


Figure 3. *Maximum head velocities obtained from each participant.*

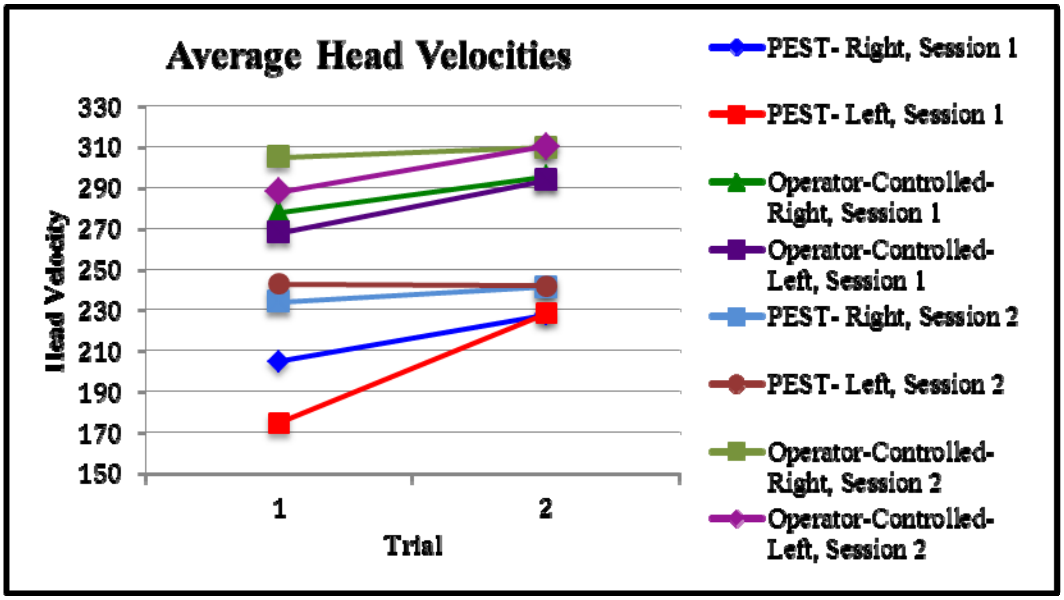


Figure 4. Average maximum head velocities achieved across testing conditions, trials, and sessions.